Optimization of an SQL Database Towards Selective Targeting of Acute Myeloid Leukemia Cells
(Technical Paper)

How Machine Learning and Artificial Intelligence Can Prevent the Next Pandemic
(STS Paper)

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Biomedical Engineering

By
Casey Evans

December 6, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Shannon Barker, PhD, Department of Biomedical Engineering

Bryn E. Seabrook, PhD, Department of Engineering and Society

Introduction

Diseases, and humanity's struggle to control them, have been a defining part of human history. Malaria, which remains one of the deadliest diseases on the planet, has been responsible for the deaths of an estimated 4 to 5 percent of all people who have ever lived (Harford, 2013). While malaria still remains a very important issue, in the future the diseases that will pose the biggest threat to humanity will come in the form of pandemics. As evidenced by the global response to COVID-19, our increasingly globalized and interconnected world makes it very hard to have a controlled response to new infectious diseases. This problem is important, considering another pandemic within our lifetimes is more likely than some people might like to admit (Marani et al., 2021). Given this challenge, we must use the best tools that we have at our disposal. Since these are problems that involve complex systems with many inputs and outputs, Artificial Intelligence (AI) and machine learning (ML) technologies have the potential to be very helpful. The proposed STS research paper will examine how humanity can use AI and Machine learning to prevent and minimize future pandemics.

Aside from pandemics, it is also important to consider other, non-infectious diseases and how we can prevent the harm they cause. For example, Acute Myeloid Leukemia (AML) is the most prevalent form of leukemia in adults and causes in excess of 10,000 deaths per year (De Kouchkovsky, et al.). In addition to examining how we can use AI and machine learning to reduce the harm caused by pandemics, the proposed technical research will examine protein targets for new AML treatments and redesigning the database that helps discover these targets.

Technical Topic

Acute Myeloid Leukemia (AML) is a disease which occurs when there is an accumulation of poorly differentiated cells in the blood and bone marrow (Meyers, et al., 2013).

Despite recent advances in treatment, it is estimated that up to 70% of AML patients 65 years or older will die as a result of the disease within 1 year of their diagnosis (Döhner et al., 2015). Improving the standard of care therapy for AML could save thousands of lives each year (in the United States alone) and would provide the hope that patients desperately need.

ZielBio, a young pharmaceutical company located in Charlottesville, Virginia aims to find treatments for deadly diseases by reversing the typical discovery process. Their discovery process works by conducting hundreds of screens in order to identify potential targets that could be used in the creation of a treatment (Brinton, 2016). A screen, in this case, is when disease cells are exposed to a large number of peptide sequences in order to examine binding behavior. A target is a protein that is potentially relevant to the process of how the disease works, and therefore has the potential to be the foundation of a successful treatment. Through this process ZielBio is able to discover novel targets that have not yet been identified by other discovery methods.

In the proposed technical project, a relevant target for AML will be identified through this process using MOLM-13 cells. These cells were taken (with informed consent) from a man with AML in 1995 and have been immortalized into a cell line, which allows for them to be used for research purposes across the world (Matsuo et al., 1997). Once a target has been identified, it will be validated through a series of biological tests to confirm its identity, and additional tests will be performed to assess the nature of the relationship between the target protein and the cells of interest.

In addition to the identification of a novel AML target, this project will involve a redesign of the database used within the drug discovery platform. The database currently holds data from over 300 screens. However, the database was built in real time and therefore lacks

efficiency in its search processes. A major goal of the project will be to re-index the SQL database in order to improve search functionality. By improving the search functionality of the database, the ZielBio team will be able to spend less time waiting for database queries to process, and more time doing the important work they perform on a daily basis. By identifying a novel target relevant to Acute Myeloid Leukemia and redesigning the platform's database, this project will improve the likelihood of treatment developments in the coming years and hopefully result in lives being saved.

STS Topic

Society inherently requires people to gather, communicate, and interact with one another. As a result, humans have always dealt with outbreaks of highly infectious diseases, often called pandemics. Pandemics have occurred at a relatively regular frequency throughout human history, and have been responsible for millions if not billions of deaths (Marani et al., 2021). As a result, it should be a top priority of the human species to figure out how to minimize the harm that pandemics cause, given that they are nearly certain to continue in the future. Given their ability to handle large swaths of information and generate useful predictions and insights, Machine Learning and Artificial Intelligence can likely help reduce the severity of future pandemics, or even help prevent them entirely.

Machine Learning is a modeling strategy that involves using collected data to predict future outcomes through the use of tuning parameters and numerical optimization techniques (Bzdok et al., 2017). For biological phenomena that do not have established models, such as the spreading of diseases that are not yet fully understood by the scientific community, machine learning can be a powerful tool for identifying patterns and generating insight.

Artificial Intelligence refers to a broad range of fields that all involve using and creating software that can learn to recognize patterns in data through iterative processing and specialized algorithms (SAS India). Machine learning is perhaps the most prominent form of AI, but it is simply one method within a much larger field that includes technologies such as neural networks, natural language processing, and other subfields. In addition to using machine learning to predict future outbreaks and how diseases will spread, AI will be useful in other ways, such as improving security and screening protocols among nucleic acid and peptide synthesis providers to prevent a manmade pandemic from occurring (Lee, 2019).

This issue is fairly unique in that all of humanity has a stake in its outcome. Just as millions of people were infected with COVID-19 as a result of early failures to prevent its spread, a future pandemic could impact the entire planet if we fail to act appropriately (Katella, 2021). As far as groups are concerned, government organizations focused on public health and disease (such as the CDC, NIH, European Centre for Disease Prevention and Control, etc.) are primary stakeholders involved in these issues and will likely implement some of the strategies discussed.

It will also be important to consider the role of these technologies in broader society and how their development brought change or was the result of change. In the same vein as described by Winner in “Do Artifacts Have Politics,” AI can be said to have an inherent political value because it requires input from society in the form of data, and this data must have a source (Winner, 1980). The system cannot exist in a vacuum outside of societal influence. Given that our discussion is centered around communicable diseases between humans, this data will be that of individuals or groups of people, and therefore there must be some value judgments made surrounding privacy. Critics of the concept of political technologies would argue that humans

simply apply value to technology, and that it is not inherently political. In addition to considering the political nature of these topics, it will also be important to acknowledge that decisions in the realm of public health are always going to involve a "calculus of risk" as discussed by Gabe Mythen (Mythen, 2004). Typically, a healthy society creates a balance of risk and freedom, as a maximally free society (no order) and a maximally safe society (no freedom) would both be states of misery for everyone involved. These tradeoffs and risk analysis must be considered carefully when thinking about initiatives that may be taken to combat the spread of a pandemic (Miller, 2019). Critics of risk analysis would argue that it does not take social dynamics and emotions into account and simply reduces people and their lives down to numbers and data. In order to account for this critique, the proposed STS research paper will examine how to balance the social value of community and connection with the need to prevent unneeded tragedy through the spread of pandemics. For example, if an AI system predicts that a pandemic will soon rise out of a certain region, a rational decision must be made using the probability of this event occurring combined with the negative effect of a given solution (such as shutting down all social/in-person activity in that region).

Research Question and Methods

In the forthcoming paper I will explore how we as a society can use machine learning and artificial intelligence to minimize the negative effects of pandemics. The research needed will be performed by searching for academic papers using keywords and key terms such as "AI," "Pandemic," "COVID-19," "Machine Learning," "Disease X," and other related terms. These terms were selected because they are highly relevant to the topic at hand. Specifically regarding "Disease X", this is a term coined by the World Health Organization to describe a currently unknown future deadly disease, and it has been adopted by many academics and officials in

discussions of how we can prepare for pandemics (The Economist, 2018). This method of research is aligned very closely with the topic at hand, as many of the techniques in this area of research have not yet been implemented, and still remain in the realm of theory and academic discussion.

Conclusion

The technical portion of this project will identify a novel AML target while improving the efficiency of the drug discovery platform's database through reindexing. Promising results from this project could eventually lead to clinical treatments and help increase the chance of survival for AML patients. Additionally, improving the functionality of the platform's database will help ZielBio discover novel targets and pursue potential treatments for other serious diseases in addition to AML. The STS portion of this project will identify practical ideas regarding how governments and other stakeholders can use Machine Learning and Artificial Intelligence to reduce the negative outcomes associated with future pandemics. Doing so will provide a summary as to how we can be prepared for when the next pandemic eventually occurs, and encourage those in power to take action in accordance with what has been outlined. Both projects will contribute to the body of knowledge surrounding deadly diseases, with the hope that eventually we find cures and prevention methods to nullify their effects.

References

Artificial Intelligence (AI) – what it is and why it matters. SAS India. (n.d.). Retrieved November 1, 2022, from https://www.sas.com/en_in/insights/analytics/what-is-artificial-intelligence.html

Bzdok, D., Krzywinski, M., & Altman, N. (2017). Points of Significance: Machine learning: a primer. *Nature methods*, *14*(12), 1119–1120. https://doi.org/10.1038/nmeth.4526

Brinton LT, Bauknight DK, Dasa SSK, Kelly KA (2016) PHASTpep: Analysis Software for Discovery of Cell-Selective Peptides via Phage Display and Next-Generation Sequencing. PLoS ONE 11(5): e0155244. https://doi.org/10.1371/journal.pone.0155244

De Kouchkovsky, I., & Abdul-Hay, M. (2016). 'Acute myeloid leukemia: a comprehensive review and 2016 update'. *Blood cancer journal*, *6*(7), e441. https://doi.org/10.1038/bcj.2016.50

Döhner, H., Weisdorf, D. J., & Bloomfield, C. D. (2015). Acute Myeloid Leukemia. *New England Journal of Medicine*, *373*(12), 1136–1152. https://doi.org/10.1056/NEJMra1406184

Harford, T. (Host). (2013, October 5). Have Mosquitoes Killed Half the World? [Audio podcast episode]. In *More or Less*. BBC. https://www.bbc.co.uk/sounds/play/p01hnv98

Katella, K. (2021, March 9). *Our Pandemic Year - A COVID-19 timeline*. Yale Medicine. Retrieved November 2, 2022, from https://www.yalemedicine.org/news/covid-timeline

Lee, Y.-C. J., Cowan, A., & Tankard, A. (2022). Peptide Toxins as Biothreats and the Potential for AI Systems to Enhance Biosecurity. *Frontiers in Bioengineering and Biotechnology*, *10*. https://doi.org/10.3389/fbioe.2022.860390

Marani, M., Katul, G. G., Pan, W. K., & Parolari, A. J. (2021). Intensity and frequency of extreme novel epidemics. Proceedings of the National Academy of Sciences, 118(35), e2105482118. https://doi.org/10.1073/pnas.2105482118

Matsuo, Y., MacLeod, R., Uphoff, C., Drexler, H., Nishizaki, C., Katayama, Y., Kimura, G., Fujii, N., Omoto, E., Harada, M., & Orita, K. (1997). Two acute monocytic leukemia (AML-M5a) cell lines (MOLM-13 and MOLM-14) with interclonal phenotypic heterogeneity showing MLL-AF9 fusion resulting from an occult chromosome insertion, ins(11;9)(q23;p22p23). *Leukemia*, *11*(9), 1469–1477. https://doi.org/10.1038/sj.leu.2400768

Meyers, J., Yu, Y., Kaye, J. A., & Davis, K. L. (2013). Medicare fee-for-service enrollees with primary acute myeloid leukemia: an analysis of treatment patterns, survival, and healthcare resource utilization and costs. *Applied health economics and health policy*, *11*(3), 275–286. https://doi.org/10.1007/s40258-013-0032-2

Miller, S. (2019). Machine Learning, Ethics and Law. *Australasian Journal of Information Systems*, *23*. https://doi.org/10.3127/ajis.v23i0.1893

Mythen, G. (2004). Defining Risk. *Ulrich Beck: A Critical Introduction to the Risk Society*. (pp. 53-73). London, England. Sterling, Virginia. Pluto Press.

What is disease X? The Economist. (2018, March 23). Retrieved November 4, 2022, from https://www.economist.com/the-economist-explains/2018/03/23/what-is-disease-x

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1,), 121–136.